

The κ -generalized distribution: A new descriptive model for the size distribution of incomes

F. Clementi ^{a,*}, T. Di Matteo ^b, M. Gallegati ^a, G. Kaniadakis ^c

^a*Department of Economics, Polytechnic University of Marche, Piazzale R. Martelli 8, 60121 Ancona, Italy*

^b*Applied Mathematics, Research School of Physical Sciences and Engineering, The Australian National University, 0200 Canberra, Australia*

^c*Department of Physics, Polytechnic University of Turin, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Abstract

This paper proposes the κ -generalized distribution as a model for describing the distribution and dispersion of income within a population. Formulas for the shape, moments and standard tools for inequality measurement—such as the Lorenz curve and the Gini coefficient—are given. A method for parameter estimation is also discussed. The model is shown to fit extremely well the data on personal income distribution in Australia and the United States.

Key words: Income distribution, income inequality, κ -generalized statistics

PACS: 02.50.Ng, 02.60.Ed, 89.65.Gh

1 Introduction

In the analysis of income distributions, analysts have found it useful to have distributional summaries based on estimates of specific parametric functional forms, not only for their suitability in modelling some features of many empirical income distributions, but also because of their role as equilibrium distributions in economic processes [1].

* Corresponding author: Tel.: +39-071-22-07-103; fax: +39-071-22-07-102.

Email addresses: fabio.clementi@univpm.it (F. Clementi),
tiziana.dimatteo@anu.edu.au (T. Di Matteo), mauro.gallegati@univpm.it
(M. Gallegati), giorgio.kaniadakis@polito.it (G. Kaniadakis).

Vilfredo Pareto first proposed a model of income distribution in the form of a probability density function in 1897 [2], providing a description of the density for income values above some lower bound, $x_0 > 0$ [3]. If one focuses on the distribution amongst those with income greater than x_0 , there are simple expressions for the moments which depend only on the Pareto parameters α and x_0 . Moreover, the expressions for most common inequality measures depend only on α , so that the (inverse of) α may also be considered as an inequality measure.

However, the apparent attractions of the Pareto distribution evaporate somewhat when one considers its implications for the distribution of income amongst the population as a whole, *i.e.* including units with income less than x_0 . For example, Ref. [4] shows how the expression for the Gini coefficient depends on assumptions about the size of excluded population (*i.e.* the proportion of the population with income below x_0) and its average income. In particular, α no longer has such a straightforward interpretation. For example, an increase in α may be associated with a decrease in inequality according to the Gini coefficient, but an increase according to the coefficient of variation.

Later empirical studies showed that the income distribution is right-skewed and has a fat right-hand tail, and that the Pareto distribution accurately models only high levels of income, but does a poor job in describing the lower end of the distribution. As research continued, new models were proposed that better describe the data, using both a combination of known statistical distributions [5] and parametric functional forms for the distribution of income as a whole, including two-parameter models such as the lognormal and gamma, three-parameter distributions such as the Singh-Maddala and Dagum I, and four-parameter distributions such as the generalized beta distributions of the first and second kind (see the comprehensive survey by Ref. [6]).

The Pareto fat tails was observed experimentally also in physical statistical systems, and are located in the high energy region where the laws of classical physics are replaced by the relativistic ones. After 2001, a physical mechanism emerging in the context of special relativity was proposed by one of us [7], predicting a deformation of the exponential function. According to this mechanism, the classical exponential distribution transforms into a new distribution, which at high energies presents a Pareto fat tail. More precisely, this mechanism deforms the ordinary exponential function $\exp(x)$ into the generalized exponential function $\exp_\kappa(x)$ given by

$$\exp_\kappa(x) = \left(\sqrt{1 + \kappa^2 x^2} + \kappa x \right)^{\frac{1}{\kappa}}. \quad (1)$$

The above deformation is generated by the fact that the propagation of the information has a finite speed, and the deformation parameter κ is proportional to the reciprocal of this speed. The κ -generalized exponential has the

important properties

$$\exp_{\kappa}(x) \underset{x \rightarrow \pm\infty}{\sim} |2\kappa x|^{\pm \frac{1}{|\kappa|}}, \quad (2a)$$

$$\exp_{\kappa}(x) \underset{x \rightarrow 0}{\sim} \exp(x). \quad (2b)$$

It is remarkable that for classical systems where the information propagates instantaneously it results $\kappa = 0$, so that the ordinary exponential emerges naturally after noting that $\exp_0(x) = \exp(x)$. Moreover, in the low energy region $x \rightarrow 0$ according to Eq. (2b) the exponential distribution emerges again, because the system behaves classically. On the contrary, in systems where the information propagates with a finite speed—these systems are intrinsically relativistic—it results $\kappa \neq 0$, so that the exponential tails become fat according to Eq. (2a) and the Pareto law emerges.

The generalized exponential represents a very useful and powerful tool to formulate a new statistical theory capable to treat systems described by distribution functions exhibiting power-law tails and admitting a stable entropy [8,9]. Furthermore, non-linear evolution models already known in statistical physics [10] can be easily adapted or generalized within the new theory.

After 2001, the function $\exp_{\kappa}(x)$ was adopted successfully in the analysis of various physical and non physical systems. In Ref. [11] we have used the function $\exp_{\kappa}(x)$ to model the personal income distribution by defining the Complementary Cumulative Distribution Function (CCDF) through

$$P_{>}(x) = \exp_{\kappa}(-\beta x^{\alpha}), \quad x \in \mathbf{R}^{+}, \quad \alpha, \beta > 0, \quad \kappa \in [0, 1), \quad (3)$$

where the income variable x is defined as $x = \frac{z}{\langle z \rangle}$, being z the absolute personal income and $\langle z \rangle$ its mean value. The corresponding Probability Density Function (PDF) reads

$$p(x) = \frac{\alpha \beta x^{\alpha-1} \exp_{\kappa}(-\beta x^{\alpha})}{\sqrt{1 + \kappa^2 \beta^2 x^{2\alpha}}}. \quad (4)$$

Follows immediately that in this model for low incomes the CCDF behaves as a stretched exponential $P_{>}(x) = \exp(-\beta x^{\alpha})$, while at high incomes follows the Pareto law $P_{>}(x) = (2\beta\kappa)^{-\frac{1}{\kappa}} x^{-\frac{\alpha}{\kappa}}$. Similarly, the PDF for $x \rightarrow 0^{+}$ behaves as a Weibull distribution $p(x) = \alpha \beta x^{\alpha-1} \exp(-\beta x^{\alpha})$, while for $x \rightarrow +\infty$ reduces to the Pareto's law $p(x) = \frac{\alpha}{\kappa} (2\beta\kappa)^{-\frac{1}{\kappa}} x^{-(\frac{\alpha}{\kappa}+1)}$.

Starting from the definitions in Eqs. (3) and (4), in this work we derive the basic statistical properties of the proposed distribution along with common tools that are required for income distribution analysis; these include, among others, the ubiquitous Lorenz curve and the associated Gini measure of inequality. The basic proposition of this paper is that the κ -distribution provides a very

good description of the size distribution of income, ranging from the low region to the middle region, and up to the power-law tail, and the inequality analysis expressed in terms of its parameters reveals very powerful.

The content of the paper is organized as follows: Sec. 2 includes a discussion of the κ -distribution and reports formulas which are useful in the estimation and analysis of empirical data. Sec. 3 illustrates applications of the results to Australian and US household survey data. Sec. 4 concludes.

2 The κ -generalized distribution

2.1 Basic properties

Using the complementary relation $P_{\leq}(x) = 1 - P_{>}(x)$, we see that the quantile function is available in closed form

$$x = P_{\leq}^{-1}(u) = \beta^{-\frac{1}{\alpha}} \left[\log_{\kappa} \left(\frac{1}{1-u} \right) \right]^{\frac{1}{\alpha}}, \quad 0 < u < 1, \quad (5)$$

a property that facilitates the derivation of Lorenz-ordering results (see Sec. 2.3). From Eq. (5) we easily determine that the median of the distribution is $x_{\text{med}} = \beta^{-\frac{1}{\alpha}} [\log_{\kappa}(2)]^{\frac{1}{\alpha}}$.

The mode is at

$$x_{\text{mode}} = \beta^{-\frac{1}{\alpha}} \left\{ \left[\frac{\alpha + 2\kappa^2(\alpha - 1)}{2\kappa^2(\alpha^2 - \kappa^2)} \right] \left(\sqrt{1 + \frac{4\kappa^2(\alpha^2 - \kappa^2)(\alpha - 1)^2}{[\alpha^2 + 2\kappa^2(\alpha - 1)]^2}} - 1 \right) \right\}^{\frac{1}{2\alpha}} \quad (6)$$

if $\alpha > 1$; otherwise, the distribution is zero-modal with a pole at the origin.

2.2 Moments and related parameters

The r^{th} -order moment about the origin of the κ -generalized distribution equals

$$\mu'_r = \int_0^{\infty} x^r p(x) dx = \frac{(2\beta\kappa)^{-\frac{r}{\alpha}} \Gamma\left(\frac{1}{2\kappa} - \frac{r}{2\alpha}\right)}{1 + \frac{r}{\alpha}\kappa} \frac{\Gamma\left(\frac{1}{2\kappa} + \frac{r}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa} + \frac{r}{2\alpha}\right)} \Gamma\left(1 + \frac{r}{\alpha}\right), \quad (7)$$

where $\Gamma(x)$ is the Gamma function $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$. Specifically, $\mu'_1 = m$ is the mean of the distribution.

A formula for the variance is obtained by converting Eq. (7) to the moment about the mean using the general equation $\mu_r = \sum_{j=0}^r \binom{r}{j} (-1)^{r-j} \mu'_j m^{r-j}$; hence, for $r = 2$ we have

$$\sigma^2 = (2\beta\kappa)^{-\frac{2}{\alpha}} \left\{ \frac{\Gamma\left(1 + \frac{2}{\alpha}\right) \Gamma\left(\frac{1}{2\kappa} - \frac{1}{\alpha}\right)}{1 + 2\frac{\kappa}{\alpha}} \frac{\Gamma\left(\frac{1}{2\kappa} + \frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa} + \frac{1}{\alpha}\right)} - \left[\frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \Gamma\left(\frac{1}{2\kappa} - \frac{1}{2\alpha}\right)}{1 + \frac{\kappa}{\alpha}} \frac{\Gamma\left(\frac{1}{2\kappa} + \frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa} + \frac{1}{2\alpha}\right)} \right]^2 \right\}. \quad (8)$$

In this way it is also possible to define the standardized moments of the distribution, which are in turn used to define skewness and excess kurtosis, respectively given by

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu'_3 - 3m\sigma^2 - m^3}{\sigma^3} \quad (9)$$

and

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu'_4 - 3\sigma^4 - 4\gamma_1\sigma^3m - 6\sigma^2m^2 - m^4}{\sigma^4}. \quad (10)$$

2.3 Lorenz curves and inequality measures

For a discussion of income inequality, the standard practice adopts the concept of concentration of incomes as defined by Lorenz [12]. The so-called Lorenz curve measures the cumulative fraction of population with incomes below x along the horizontal axis, and the fraction of the total income this population accounts for along the vertical axis. The points plotted for the various values of x trace out a curve below the 45° line sloping upwards to the right from the origin.

In statistical terms, for any general distribution supported on the nonnegative half-line with a finite and positive first moment the Lorenz curve can be written in the form $L(u) = \int_0^u P_{\leq}^{-1}(t) dt / \int_0^1 P_{\leq}^{-1}(u) du$, $u \in [0, 1]$, where $m = \int_0^1 P_{\leq}^{-1}(u) du$ is the quantile formula for the mean, and $P_{\leq}^{-1}(u)$ is the quantile function given by Eq. (5)¹. Thus, we have the Lorenz curve for the κ -generalized distribution as follows

$$L_\kappa(u) = 1 - \frac{1 + \frac{\kappa}{\alpha}}{2\Gamma\left(\frac{1}{\alpha}\right)} \frac{\Gamma\left(\frac{1}{2\kappa} + \frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa} - \frac{1}{2\alpha}\right)} \left\{ 2\alpha(2\kappa)^{\frac{1}{\alpha}} (1-u) \left[\log_\kappa \left(\frac{1}{1-u} \right) \right]^{\frac{1}{\alpha}} + \right. \\ \left. + B_X \left(\frac{1}{2\kappa} - \frac{1}{2\alpha}, \frac{1}{\alpha} \right) + B_X \left(\frac{1}{2\kappa} - \frac{1}{2\alpha} + 1, \frac{1}{\alpha} \right) \right\}, \quad (11)$$

where $B_x(s, r)$ is the incomplete Beta function given by $B_x(s, r) = \int_0^x t^{s-1} (1-t)^{r-1} dt$ with $X = (1-u)^{2\kappa}$.

¹ See *e.g.* Ref. [13].

The Gini measure of income inequality [14] can be derived using the representation $G = 1 - \frac{1}{m} \int_0^\infty [1 - P_\leq(x)]^2 dx$ given by Ref. [15]; it follows that the Gini coefficient for the κ -generalized distribution is

$$G_\kappa = 1 - \frac{2\alpha + 2\kappa}{2\alpha + \kappa} \frac{\Gamma\left(\frac{1}{\kappa} - \frac{1}{2\alpha}\right) \Gamma\left(\frac{1}{2\kappa} + \frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{\kappa} + \frac{1}{2\alpha}\right) \Gamma\left(\frac{1}{2\kappa} - \frac{1}{2\alpha}\right)}. \quad (12)$$

Furthermore, relating the standard deviation to the mean yields the following expression for the coefficient of variation

$$CV_\kappa = \frac{\sigma}{m} = \sqrt{\frac{\frac{\Gamma(1+\frac{2}{\alpha})}{1+2\frac{\kappa}{\alpha}} \frac{\Gamma(\frac{1}{2\kappa}-\frac{1}{\alpha})}{\Gamma(\frac{1}{2\kappa}+\frac{1}{\alpha})}}{\left[\frac{\Gamma(1+\frac{1}{\alpha})}{1+\frac{\kappa}{\alpha}} \frac{\Gamma(\frac{1}{2\kappa}-\frac{1}{2\alpha})}{\Gamma(\frac{1}{2\kappa}+\frac{1}{2\alpha})}\right]^2}} - 1. \quad (13)$$

2.4 Estimation

Parameter estimation for the κ -generalized distribution can be performed using the Maximum Likelihood (ML) approach. Assuming that all observations $\mathbf{x} = \{x_1, \dots, x_n\}$ are independent, the likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n p(x_i) = (\alpha\beta)^n \prod_{i=1}^n \frac{x_i^{\alpha-1} \exp_\kappa(-\beta x_i^\alpha)}{\sqrt{1 + \beta^2 \kappa^2 x_i^{2\alpha}}}, \quad (14)$$

where $\boldsymbol{\theta} = \{\alpha, \beta, \kappa\}$ is the parameter vector. This leads to the problem of solving the partial derivatives of the log-likelihood function $l(\boldsymbol{\theta}; \mathbf{x}) = \log L(\boldsymbol{\theta}; \mathbf{x})$ with respect to κ , α and β . However, obtaining explicit expressions for the ML estimators of the three parameters is difficult, making direct analytical solutions intractable, and one needs to use numerical optimization methods.

Taking into account the meaning of the variable x , the mean value results to be equal to unity, *i.e.* $m = \int_0^\infty xp(x) dx = 1$. The latter relationship permits to express the parameter β as a function of the parameters κ and α , obtaining

$$\beta = \frac{1}{2\kappa} \left[\frac{\Gamma\left(\frac{1}{\alpha}\right) \Gamma\left(\frac{1}{2\kappa} - \frac{1}{2\alpha}\right)}{\kappa + \alpha \Gamma\left(\frac{1}{2\kappa} + \frac{1}{2\alpha}\right)} \right]^\alpha. \quad (15)$$

In this way, the problem to determine the values of the free parameters $\{\kappa, \alpha, \beta\}$ of the theory from the empirical data reduces to a two parameter $\{\kappa, \alpha\}$ fitting problem. Therefore, to find the parameter values that give the most desirable fit, one can use the Constrained Maximum Likelihood (CML) estimation method [16], which solves the general maximum log-likelihood

problem of the form $l(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^n \log p(x_i; \boldsymbol{\theta})^{w_i}$, where n is the number of observations, w_i the weight assigned to each observation, $p(x_i; \boldsymbol{\theta})$ the probability of x_i given $\boldsymbol{\theta}$, subject to the non-linear equality constraint given by Eq. (15) and bounds $\alpha, \beta > 0$ and $\kappa \in [0, 1)$. The CML procedure finds values for the parameters in $\boldsymbol{\theta}$ such that $l(\boldsymbol{\theta}; \mathbf{x})$ is maximized using the sequential quadratic programming method [17] as implemented, *e.g.*, in MATLAB[®] 7.

3 Empirical application to income data

The κ -generalized distribution was fit to data on personal income distribution for Australia and the United States². The data are derived from panel surveys conducted in 2002–03 and 2003, respectively. The unit of assessment is the household, and income is expressed in nominal local currency units (and is equivalized for differences in household size by adjusting by the square root of the number of household members [18]). There are 10,211 households in the 2002–03 Australian survey, and 7,822 in the 2003 US survey. All calculations use the sampling weights produced by the data provider [19]. We consider the distributions of disposable income, *i.e.* the income recorded after the payment of taxes and government transfers. In the data analysis, we exclude the observations with zero and negative values, and normalize income to its empirical average, given by $32,891.17 \pm 343.58$ AUD and $31,812.39 \pm 598.74$ USD respectively³.

Maximum likelihood estimates are shown in panels (a) and (b) of Figs. 1 and 2. All the parameters were very precisely estimated, and the comparison between the fitted and sample estimates of the CCDF and PDF suggests that the κ -generalized distribution offers a great potential for describing the data over their whole range, from the low to medium income region through to the high income Pareto power-law regime, including the intermediate region for which a clear deviation exists when two different curves are used.

Panel (c) of the same figures depicts the data points for the empirical Lorenz curve, *i.e.* $L\left(\frac{i}{n}\right) = \sum_{j=1}^i x_j / \sum_{j=1}^n x_j$, $i = 1, 2, \dots, n$, superimposed by the

² These data were not studied in the previous paper [11], where the emphasis was on other countries. However, our main findings here have been applied also to the data included in that work, and the results are available from the authors upon request.

³ More detailed information on the Australian data is available on the Australian Bureau of Statistics (ABS) web site: <http://www.abs.gov.au>. For analyses referring to the same country and data source, see Refs. [20]. For the US data, see Refs. [21], or consult the following web address: <http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm>.

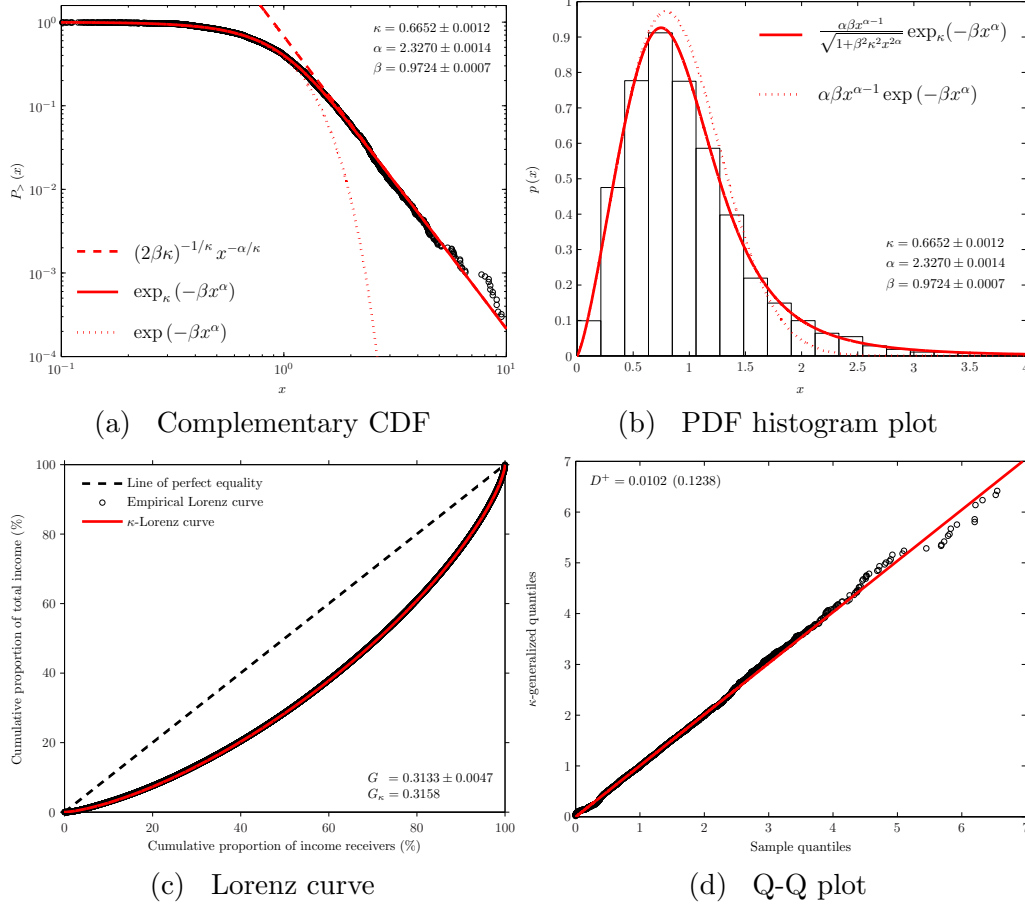


Fig. 1. The Australian personal income distribution in 2002–03 measured in current year AUD. (a) Plot of the empirical CCDF in the log-log scale. The solid line is our theoretical model given by Eq. (3) fitting very well the data in the whole range from the low to the high incomes including the intermediate income region. This function is compared with the ordinary stretched exponential one (dotted line)—fitting the low income data—and with the pure power-law (dashed line)—fitting the high income data. (b) Histogram plot of the empirical PDF with superimposed fits of the κ -generalized (solid line) and Weibull (dotted line) PDFs. (c) Plot of the Lorenz curve. The hollow circles represent the empirical data points and the solid line is the theoretical curve given by Eq. (11) using the same parameter values as in panels (a) and (b). The dashed line corresponds to the Lorenz curve of a society in which everybody receives the same income and thus serves as a benchmark case against which actual income distribution may be measured. (d) Q-Q plot of the sample quantiles versus the corresponding quantiles of the fitted κ -generalized distribution. The reference line has been obtained by locating points on the plot corresponding to around the 25th and 75th percentiles and connecting these two. In plots (a), (b) and (d) the income axis limits have been adjusted according to the range of data to shed light on the intermediate region between the bulk and the tail of the distribution.

theoretical curve $L_\kappa(u)$ given by Eq. (11) with estimates replacing α and κ as necessary. This formula is shown by the solid line in the plots, and fits the data exceptionally well. The plots also exhibit a very good agree-

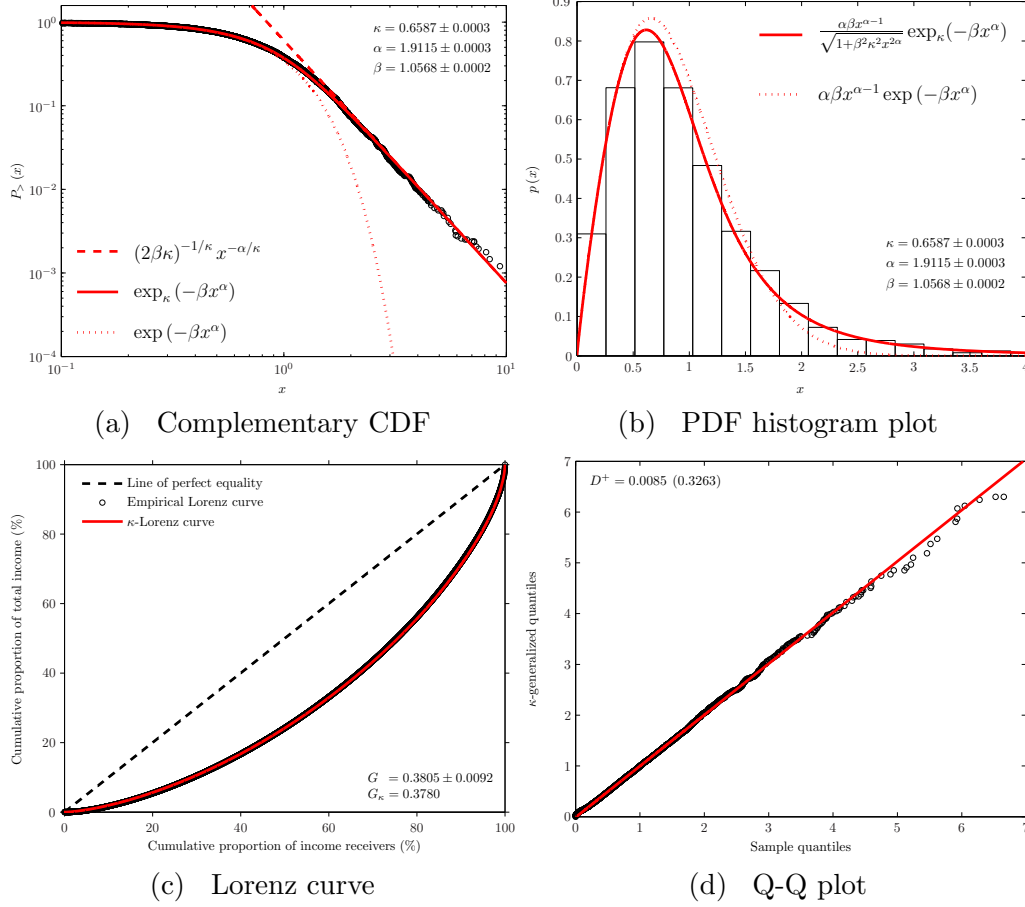


Fig. 2. Same plots as in Fig. 1 for the US personal income distribution in 2003. The income variable is measured in current year USD.

ment between the empirical estimates of the Gini coefficient, obtained as $G = \frac{1}{n^2\mu} \sum_{i=1}^n (2i - n - 1)x_i$, and the values returned by the analytical expression given by Eq. (12) for the estimated κ -generalized distribution; the 95% confidence intervals constructed around the values of G always cover the theoretical predictions G_κ ^{6, 7}.

In order to further evaluate the accuracy of our distributional model, we have also tested the hypothesis that each set of n observed data follows a κ -generalized distribution by calculating the Kolmogorov-Smirnov (K-S) goodness-of-fit test statistic given by $D^+ = \max_{1 \leq i \leq n} [in^{-1} - P_{\leq}(x_i)]$, $i = 1, 2, \dots, n$. Since in this case there is no asymptotic formula for calculating the p -value, we have reduced the problem to testing that the x values have a standard exponential distribution (*i.e.*, an exponential distribution with parameter

⁶ For the formulas used to estimate the empirical Lorenz curve and Gini coefficient see *e.g.* Refs. [22] and [23], respectively.

⁷ The confidence intervals for the observed Gini coefficients have been calculated via the bootstrap resampling method based on 1000 replications. For general details about bootstrapping, see Refs. [24].

equal to 1) by relating the function $P_{>}(x)$ given by Eq. (13) to the ordinary exponential function, namely $\exp_{\kappa}(-\beta x^{\alpha}) = \exp(-x_{\kappa})$, through the transformation $x_{\kappa} = \frac{1}{\kappa} \log(\sqrt{1 + \beta^2 \kappa^2 x^{2\alpha}} + \beta \kappa x^{\alpha})$, where the parameters are estimated from the data. Thus the significance level in the upper tail is given approximately by $P_{>}(T^*) = \exp[-2(T^*)^2]$, with $T^* = D^+(\sqrt{n} + 0.12 + 0.11/\sqrt{n})$, as suggested for example by Ref. [25]. The results are shown in the upper-left corner of panel (d) of Figs. 1 and 2. As one can appreciate, the maximum distance between the empirical data and the theoretical model as assessed by the K-S statistic is very small, and the p -values in parentheses do not lead to rejection of the null hypothesis that the data may come from a κ -generalized distribution at any of the usual significance levels (1%, 5% and 10%). The linear behaviour emerging from the Quantile-Quantile (Q-Q) plots of the sample quantiles versus the corresponding quantiles of the fitted κ -generalized distribution displayed in the same panel strongly supports the quantitative results obtained by hypothesis testing.

4 Summary and conclusions

One of the main objectives of research on income distribution is to provide a mathematical description of the size distribution of income for approximating the underlying “true” distribution. Starting from Pareto contribution, a wide variety of functional forms have been considered as possible models for the distribution of personal income by size, and other approaches can no doubt be suggested and deserve attention.

In this work we have proposed a three-parameter distribution by using a new approach having its root in the framework of the κ -generalized statistical mechanics. This model shows able to describe the entire income range, including the Pareto upper tail, and fits the Australian and US income data extremely well. The analysis of inequality performed in terms of its parameters reveals the merit of the proposed distribution, and provide the basis for a fruitful interaction between the two fields of statistical mechanics and economics.

Acknowledgements

T. Di Matteo wishes to thank the Australian Social Science Data Archive, ANU, for providing the ABS data and the partial support by ARC Discovery Projects: DP03440044 (2003) and DP0558183 (2005), COST P10 “Physics of Risk” project and M.I.U.R.-F.I.S.R. Project “Ultra-high frequency dynamics of financial markets”.

References

- [1] P. Richmond, S. Hutzler, R. Coelho, P. Repetowicz, P., in *Econophysics and Sociophysics: Trends and Perspectives*, edited by B.K. Chakrabarti, A. Chakraborti, A. Chatterjee (Wiley-VCH, Berlin, 2006), pp. 131–160.
- [2] V. Pareto, *Course d'économie politique* (Macmillan, London, 1897).
- [3] B.C. Arnold, *Pareto Distributions* (International Co-operative Publishing House, Fairland, 1983).
- [4] A.B. Atkinson, A.J. Harrison, *The Distribution of Wealth in Britain* (Cambridge University Press, Cambridge, 1978).
- [5] F. Clementi, M. Gallegati, *Physica A* **350**, 427 (2005); F. Clementi, M. Gallegati, in *Econophysics of Wealth Distributions*, edited by A. Chatterjee, S. Yarlagadda, B.K. Chakrabarti (Springer-Verlag Italia, Milan, 2005), pp. 3-14; A.A. Drăgulescu, V.M. Yakovenko, *Eur. Phys. J. B* **20**, 585 (2001); A.A. Drăgulescu, V.M. Yakovenko, *Physica A* **299**, 213 (2001); E.W. Montroll, M.F. Shlesinger, *P. Natl. Acad. Sci. USA* **79**, 3380 (1982); E.W. Montroll, M.F. Shlesinger, *J. Stat. Phys.* **32**, 209 (1983); M. Nirei, W. Souma, Two factor model of income distribution dynamics. Working paper No. 04-10-029, Santa Fe Institute (2004), available at: <http://www.santafe.edu/research/publications/wpabstract/200410029>; W. Souma, *Fractals* **9**, 463 (2001); A.C. Silva, V.M. Yakovenko, *Europhys. Lett.* **69**, 304 (2005).
- [6] C. Kleiber, S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences* (John Wiley & Sons, New York, 2003).
- [7] G. Kaniadakis, *Physica A* **296**, 405 (2001); G. Kaniadakis, *Phys. Lett. A* **288**, 283 (2001); G. Kaniadakis, *Phys. Rev. E* **66**, 056125 (2002); G. Kaniadakis, *Phys. Rev. E* **72**, 036108 (2005).
- [8] G. Kaniadakis, A.M. Scarfone, *Physica A* **340**, 102 (2004).
- [9] S. Abe, G. Kaniadakis, A.M. Scarfone, *J. Phys. A: Math. Gen.* **37**, 10513 (2004).
- [10] G. Kaniadakis, A. Lavagno, P. Quarati, *Nucl. Phys. B* **466**, 527 (1996); G. Kaniadakis, A. Lavagno, P. Quarati, *Phys. Lett. A* **227**, 227 (1997); G. Kaniadakis, P.P. Delsanto, *Math. Comput. Model.* **17**, 31 (1993).
- [11] F. Clementi, M. Gallegati, G. Kaniadakis, *Eur. Phys. J. B* **57**, 187 (2007).
- [12] M.O. Lorenz, *J. Am. Stat. Ass.* **9**, 209 (1905).
- [13] J.L. Gastwirth, *Econometrica* **39**, 1037 (1971).
- [14] C. Gini, *Metron* **63**, 3 (2005).
- [15] B.C. Arnold, L. Laguna, *On Generalized Pareto Distributions with Applications to Income Data* (Iowa State University Press, Ames, 1977).

- [16] R. Schoenberg, *Computational Econ.* **10**, 251 (1997).
- [17] S.P. Han, *J. Optimiz. Theory App.* **22**, 297 (1977).
- [18] A. Deaton, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (Johns Hopkins University Press, Baltimore MD, 1996).
- [19] Expert Group on Household Income Statistics—The Canberra Group. Final report and recommendations (2001), available at: <http://www.lisproject.org/links/canberra/finalreport.pdf>.
- [20] T. Di Matteo, T. Aste, S.T. Hyde, in *The Physics of Complex Systems (New Advances and Perspectives)*, edited by F. Mallamace, H.E. Stanley (Amsterdam: IOS Press, Amsterdam, 2004), pp. 435-442; F. Clementi, T. Di Matteo, M. Gallegati, *Physica A* **370**, 49 (2006); A. Banerjee, V.M. Yakovenko, T. Di Matteo, *Physica A* **370**, 54 (2006).
- [21] R.V. Burkhauser, B.A. Butrica, M.C. Daly, D.R. Lillard, in *Social Insurance in a Dynamic Society. Papers in Honor of the 65th Birthday of Richard Hauser*, edited by I. Becker, N. Ott, G. Rolf (Geburtstag Campus, Frankfurt and New York, 2001), pp. 354-376, available at: [http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/upload/P25-R.V.Burkhauser,P.Giles,D.R.Lillard,J.Schwarze,Creation of private and public \(social security and other government income\) categories for CNEF data. Working paper, Cornell University, Department of Policy Analysis and Management \(2004\), available at: \[www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/upload/social.pdf\]\(http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/upload/social.pdf\).](http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/upload/P25-R.V.Burkhauser,P.Giles,D.R.Lillard,J.Schwarze,Creation%20of%20private%20and%20public%20(social%20security%20and%20other%20government%20income)%20categories%20for%20CNEF%20data.pdf)
- [22] J.L. Gastwirth, *Rev. Econom. Stat.* **54**, 306 (1972).
- [23] K. Xu, How has the literature on Gini's index evolved in the past 80 years? Working paper, Dalhousie University, Department of Economics (2003), available at: <http://ssrn.com/abstract=423200>.
- [24] B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993); A.C. Davidson, D.V. Hinkley, *Bootstrap Methods and Their Application* (Cambridge University Press, New York, 1997).
- [25] M.A. Stephens, *J. Roy. Stat. Soc. B Met.* **32**, 115 (1970).